Gossip Learning under Different Network Topologies

Krystian Szczepankiewicz* Kaung Myat Naing* Zack Chaffee* Melissa Laiz*
*Department of Computer Science, Illinois Institute of Technology; Chicago, IL, United States

***Abstract*—Decentralized machine learning systems enable model training across distributed devices without centralizing raw data, addressing critical privacy-sensitive applications. However, with this network topology, the structure of communication between devices, significantly impacts learning efficiency, convergence speed, and communication overhead. This project investigates how different graph topologies affect gossip learning performance by implementing it from scratch simulation framework on Fashion-MNIST using a simple MLP architecture. Furthermore, narrowing the scope from comparing multiple decentralized paradigms (Federated Learning [FL], Hierarchical FL, and Gossip Learning) to focusing exclusively on gossip learning, we evaluate as a group performance across three fundamental random graph models. With the use of NetworkX for graph generation and metrics computation, we then measure both node-level properties and graph-level characteristics to identify further which structural features will be most influenced with the convergence behavior.**

Moreover the key contribution is evaluating different weighting schemes in the gossip averaging step. We as a group compared uniform averaging against the Metropolis-Hastings matrix weighting that adaptively weights neighbor contributions based on the nodes degrees. The overall result that will be discussed the sensitivity of network structure. The higher the average degree the faster the information mixing. Thus, demonstrates that topology and weighting schemes together determine efficiency in gossip-based decentralized learning.

***Index Terms*—Gossip Learning, Network Topology, Decentralized Machine Learning, Random Graph Models, Convergence Analysis, Peer-to-Peer Systems, Metropolis-Hastings matrix**

I. Introduction

The exponential growth of data being generated within our edge systems, 'mobile devices', 'sensors', and 'IoT systems', has made centralized machine learning impractical and often undesirable. Thus, privacy regulations and the sensitivity of personal data have motivated the development of decentralized learning paradigms that train models without collecting raw data on a central server. FL has emerged as a go-to approach, but retains a critical bottle neck: the central server that aggregates model updates. Therefore, creates a single point of failure, a privacy vulnerability, and a communication overhead caused by large model updates between numerous edge devices.

Moreover, Gossip Learning offers a peer-to-peer alternative communication to which information spread through network. To where, each node is then trained locally on its private data and exchange model updates only with its direct neighbors. Thus, information propagates through the network, like person-to-person. Therefore, this approach is decentralized, but no central server is needed, and the system is inherently robust to node failures and dynamic topology changes.

However, unlike centralized FL where all nodes communicate uniformly with a central server, gossip learning's performance is fundamentally constrained by its network topology. The structure of communication links between nodes directly determines convergence speed, information mixing rates, and overall learning efficiency. To where, a poorly connected network may suffer from slow convergence and information bottlenecks, while a well-connected topology can facilitate model consensus. Therefore, the central question this project will address is: How do structural properties of different network topologies affect gossip learning convergence under non-IID data?

So, systematically we are investigating by implementing a gossip learning framework from scratch and evaluate it across three fundamental random graph models, Erdős–Rényi (purely random), Watts–Strogatz (small-world), and Barabási–Albert (scale-free), each with distinct topological properties. Additionally, we compare two weighting schemes for neighbor averaging simple uniform weights and the Metropolis-Hastings matrix, which adaptively weight contributions based on the node degree to improve convergence to its consensus.

Thus, our evaluation will measure both graph-level metrics, diameter, average shortest path, clustering coefficients, and node-level metrics, degree, clustering, final accuracy, on Fashion-MNIST with non-IID data distribution across 30 clients. By isolating the topology and weighting scheme as key variables, we then provide the insights into which structural features will then be most influential for its convergence speed and final accuracy. This insights will be critical for designing efficient peer-to-peer machine learning systems where no central server exists.

II. Related Works

We collected, and analyzed several papers that correlates with desired topics which are the core of Gossip Learning, Network Topology and Convergence, Small-world network properties, Federated Learning & Decentralized Optimization, Gossip & Distributed Algorithms, etc.

Gossip Learning vs Federated Learning, the work by Hegedűs, Danner, and Jelasity [1], [2] demonstrates that gossip learning is viable fully decentralized alternative to federate learning, with the best gossip varieties performing comparably to federated learning overall under various scenarios including non-uniform data distributions. Moreover, their work on matrix factorization [3] further shows that in large networks, gossip learning can outperform federate learning when the same subsampling based compression techniques are applied. However, gossip learning typically suffers to slower convergence and lower final accuracy because it relies on peer-to-peer infrastructure rather than cloud resources. Moreover, McMahan et al.'s foundational FedAvg algorithm [4] established federated learning as the primary strictire, but it is reliant on a central server that motivates exploration of fully decentralized alternatives. Moreover, early work by Ormándi, Hegedűs, and Jelasity [5] proposed gossip learning as a generic approach using multiple models taking random walked over the networks in parallel, combining ensembles learning methods.

In topology and convergence, in network topology fundamentally determines the efficiency of gossip algorithms. Averaging weights particularly the Metropolis-Hastings matrix directly influences distributed consensus convergence by adaptively weighting neighbor contributions based on node degrees [6], [7]. Then the topological features including connectivity, clustering, and shortest-path distances directly influence convergence rates in distributed averaging algorithms. Jelasity et al. [8] talked about peer-sampling as first-class abstraction for decentralized systems, constructing and maintaining dynamic unstructured overlays through gossiping membership information. To where, gossip protocols are fatal for for decentralized systems because they require no specialized routing and are robust to changing topology. However, the efficiency depends critically on underlying network structure. The foundational work of Ben-Or and Upfal [9] which demonstrated randomized algorithms solve fundamental problems in distributed computing.

Lastly, the choice of network topology profoundly affects distributed learning. Foundational work introduced three canonical graph families: Erdős-Rényi (ER) [10], [11] for purely random graphs, Watts-Strogatz (WS) [12] for small-world networks, and Barabási-Albert (BA) [13] for scale-free networks with preferential attachment. Zweig [14] provides a comprehensive review of these models. Additionally, a critical challenge in decentralized learning is non-IID data across clients; while federated learning research extensively addresses this, the interaction between data heterogeneity and network topology in gossip learning remains understudied. Our work bridges this gap by empirically evaluating gossip learning under non-IID conditions (Dirichlet distribution) across diverse topologies, providing insights into how network structure affects learning dynamics in realistic heterogeneous data scenarios. Recent work by Hegedűs et al. [15] further demonstrates the viability of fully decentralized approaches.

III. System Model Description

This section describes the system used in our experiments, including the network structure, data distribution, local computation, and communication protocol. We focus on a fully decentralized gossip learning setting, where each node trains a local model on its own private data and communicates only with its direct neighbors. By keeping the model architecture simple and using static network topologies, we distinguish the effect of network structure and weighting schemes on learning behavior.

The communication network is modeled as an undirected graph $G = (V, E)$, where $|V| = N$ represents the set of nodes (clients) and E represents communication links. Each node holds local data and a copy of the model. The network topology remains fixed throughout training.

To study how topology affects gossip learning, we generate networks from three well-known random graph models. Their parameters are chosen so that different topologies have comparable average degrees k. All graphs are generated using NetworkX, which is also used to compute graph-level metrics such as average degree, clustering coefficient, diameter, and average shortest-path length.

Erdős–Rényi (ER) [10] : Edges are added independently between pairs of nodes with probability p. This results in a relatively uniform degree distribution and low clustering. ER graphs are easy to analyze and serve as a baseline random topology.

Watts–Strogatz (WS) [12] : Starting from a regular ring lattice, each edge is rewired with probability β. This model maintains high clustering while significantly reducing average path length. We fix β=0.3 and adjust the neighborhood size to match the target average degree.

Barabási–Albert (BA) [13] : Nodes are added sequentially using preferential attachment, meaning new nodes are more likely to connect to high-degree nodes. This produces a power-law degree distribution with a small number of hub nodes and many low-degree nodes.

A fixed mixing matrix $W \in \mathbb{R}^{N\times N}$ is computed once per network topology and reused throughout training. Under the Metropolis–Hastings mixing scheme, the weight between two connected nodes i and j is defined as

$$w_{ij} = \frac{1}{1+\max(d_i, d_j)}, \quad \text{for } (i,j) \in E, \tag{1}$$

where d_i and d_j denote the degrees of nodes i and j, respectively. The self-weight is chosen so that each row sums to one. These weights require only local degree information and guarantee convergence of distributed averaging.

We use the Fashion-MNIST dataset, which contains 60,000 training images and 10,000 test images across 10 classes. To model realistic data heterogeneity, the training data are split among nodes using a Dirichlet distribution with concentration parameter α. For each class, samples are divided among nodes according to proportions drawn from a Dirichlet distribution. Smaller values of α produce highly non-IID splits, where nodes focus on a few classes, while larger values approach an IID distribution.

Each node maintains an independent copy of the same multilayer perceptron (MLP) classifier. Input images are flattened into 784-dimensional vectors. The model consists of two fully connected hidden layers with ReLU activations (128 and 64 units), followed by a ten-class output layer. All nodes start from the same initial model parameters. During each training round, every node performs E local epochs of stochastic gradient descent (SGD) on its private data, producing an updated parameter vector.

After local training, nodes exchange model parameters with their neighbors. Let $\mathcal{N}(i)$ denote the set of neighbors of node i. Each node updates its model by forming a weighted average of its own parameters and those received from its neighbors:

$$\theta_i^{(t+1)} = \sum_{j \in \mathcal{N}(i) \cup \{i\}} w_{ij}\, \theta_j^{(t)}. \tag{2}$$

When using Metropolis–Hastings weights, the mixing matrix is doubly stochastic, ensuring that repeated gossip steps drive all models toward a common agreement. Communication cost is proportional to the number of exchanged parameter vectors, and we track the total bytes transmitted during training. Training proceeds for a fixed maximum number of rounds or stops early if the average test accuracy across nodes exceeds a target threshold while the accuracy variance falls below a dispersion threshold.

IV. Formal Problem Definition

We consider a fully decentralized learning system consisting of multiple nodes connected by a fixed communication network. Each node holds its own local dataset and maintains a local copy of the model. The data across nodes may be non-identically distributed, reflecting realistic heterogeneity in decentralized environments. Nodes do not share raw data and there is no central server coordinating training.

Our objective is to collaboratively learn a model that performs well across the overall data distribution. Since no node has access to the complete dataset, this objective is achieved through local training combined with peer-to-peer communication. Each node improves its model using its private data and periodically exchanges model parameters with neighboring nodes, allowing information to gradually propagate through the network.

The underlying network topology limits peer-to-peer communication. Nodes are allowed to communicate only with their immediate neighbors, and no global knowledge of the network is assumed. Model updates rely only on locally available information, ensuring that the learning process remains fully decentralized and scalable.

The learning process is evaluated based on two main criteria. First, consensus, which measures whether the models across different nodes become increasingly similar over time. Second, accuracy, which measures how well the learned model performs on unseen test data. In this work, we study how different network topologies and averaging strategies influence both the speed of convergence toward consensus and the final predictive performance, particularly in the presence of non-IID data distributions.

V. Algorithm Design

Algorithm 1 Gossip Learning

for each round $t = 1$ to T **do**
  for each node i **in parallel do**
    Train local model on its own data D_i for a few epochs
    $\theta_i \leftarrow \theta_i - \eta \nabla \mathcal{L}(\theta_i; D_i)$
  end for
  for each node i **in parallel do**
    Send current model θ_i to all neighbors
    Receive models θ_j from all neighbors j
  end for
  for each node i **in parallel do**
    Average all received models (including own):
    $\theta_i \leftarrow \sum_{j \in \{i\} \cup \mathcal{N}(i)} w_{ij}\theta_j$ $\{w_{ij}$ = weights from Metropolis-Hastings$\}$
  end for
end for=0

VI. Methodology

We follow the system model and gossip protocol described in Section III and focus here on the experimental configuration used to evaluate topology-dependent convergence behavior. Specifically, this section describes the experimental setup and procedures used to study how different communication topologies influence convergence in gossip learning. Our objective is to systematically compare network structures while controlling for connectivity and ensuring fair and reproducible experimentation. All experiments were implemented in PyTorch, and the full codebase is publicly available on GitHub.

A. Data and Partitioning

We use the full Fashion-MNIST dataset, which consists of 60,000 training samples and 10,000 test samples across 10 classes. To emulate realistic decentralized learning scenarios, the training dataset is partitioned across clients using a non-IID Dirichlet distribution with concentration parameter α. This approach introduces heterogeneity in local data distributions while ensuring that the union of all client datasets exactly covers the full training set.

Formally, for each class c, proportions are sampled from a Dirichlet distribution and used to assign class-specific samples to clients. This method produces skewed local class distributions while preserving full dataset coverage within a single experiment.

B. Graph Topologies

We evaluate three canonical random graph models that represent different structural properties:

- **Erdős–Rényi (ER):** A random graph where each edge exists independently with probability $p = \langle k \rangle/(n-1)$, resulting in a homogeneous degree distribution.
- **Watts–Strogatz (WS):** A small-world graph constructed using a fixed degree $k = \langle k \rangle$ and a rewiring probability p_r, yielding high clustering and short average path lengths.

- **Barabási–Albert (BA):** A scale-free graph generated via preferential attachment with parameter $m \approx \langle k \rangle/2$, producing a heavy-tailed degree distribution with hub nodes.

For each experiment, we enforce that the expected average degree $\langle k \rangle$ is matched across all three graph models. This design isolates the impact of topology from differences in overall connectivity.

C. Gossip Learning Protocol

Each client maintains a local neural network with identical architecture, consisting of a three-layer multilayer perceptron. Training proceeds in synchronized gossip rounds. During each round, clients perform local stochastic gradient descent (SGD) on their private data, followed by peer-to-peer model averaging with neighboring clients defined by the communication graph.

Model aggregation is performed using a Metropolis–Hastings mixing matrix, which ensures degree-aware and doubly stochastic updates. For clients i and j with degrees d_i and d_j, the mixing weights are defined as

$$W_{ij} = \frac{1}{1 + \max(d_i, d_j)}, \quad W_{ii} = 1 - \sum_{j \neq i} W_{ij}. \tag{3}$$

This formulation prevents high-degree nodes from dominating the aggregation process and guarantees stable convergence.

D. Experimental Parameters

Unless otherwise stated, all reported experiments are conducted using networks of 30 clients, a learning rate of 0.01, and two local training epochs per gossip round. Each configuration is evaluated across multiple random seeds to account for variability in graph generation and data partitioning.

Earlier stages of this project explored smaller networks (e.g., 12 clients) to validate correctness, debug implementation details, and iterate on experimental design. The final results presented in this paper focus on 30-node networks as a balance between computational feasibility and sufficiently rich topological behavior. This progression reflects a common research workflow, moving from small-scale prototyping to larger, more representative experimental settings.

Convergence is defined using a dual criterion: (1) the mean classification accuracy across all clients exceeds 70%, and (2) the standard deviation of client accuracies is below 5%. This ensures that both model performance and consensus across clients are achieved.

We record per-round statistics including mean accuracy, dispersion, communication cost (measured in megabytes of transmitted parameters), and runtime. In addition, we compute graph-level metrics such as average degree, clustering coefficient, diameter, and average shortest path length, as well as node-level metrics including degree and local clustering.

VII. Results

This section presents the experimental results from gossip learning experiments conducted on networks of 30 nodes using three different network topologies: Erdős-Rényi (ER), Watts-Strogatz (WS), and Barabási-Albert (BA). Each topology was evaluated under two mixing strategies: Metropolis-Hastings and Uniform mixing. All experiments were trained for a maximum of 40 rounds with 2 local epochs per round using a learning rate of 0.01 on the Fashion-MNIST dataset with non-IID data distribution (Dirichlet $\alpha = 1.0$). The target average degree was set to $k = 4$ for all networks. The following figures analyze convergence behavior, final accuracy, and accuracy dispersion across these experimental configurations.

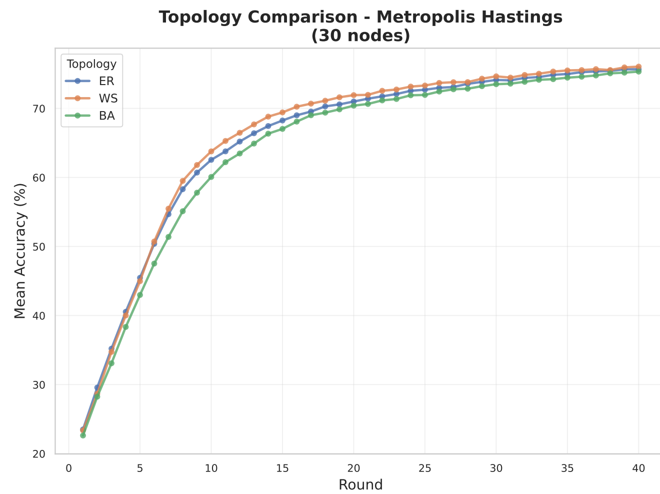

Fig. 1. Average accuracy over rounds using Metropolis-Hastings mixing.

Figure 1 compares the three topologies in terms of their accuracy over 40 rounds using Metropolis-Hastings mixing. All three topologies start around 24-25% accuracy and converge to the range of 71.90% to 73.50%. ER and WS converged slightly faster, reaching convergence at rounds 18 and 16 respectively, whilst BA took until round 20 to converge. The trend can be seen within all topologies of rapid improvement in the first 15 rounds followed by a steady plateauing, which is typical of gossip learning as model weights propagate through the network and local training updates become incremental.

Similar trends can be seen in Figure 2 with Uniform mixing as existed with Metropolis-Hastings. The accuracies start at similar levels (around 24%) and follow comparable learning curves. However, Uniform mixing achieved slightly higher final accuracies across all topologies of 74.11% for ER, 74.02% for WS, and 73.73% for BA. BA shows notably improved convergence speed with Uniform mixing (round 16) compared to Metropolis-Hastings (round 20), and overall there is less variance in convergence times between the different topologies. This suggests that Uniform mixing may be more robust to network structure variations.

When directly comparing Metropolis-Hastings and Uniform mixing as seen in Figure 3, similar overall behavior is observed between the two strategies. Both exhibit a gradual rise in accuracy until reaching a plateau around rounds 15-20. Uniform mixing achieves approximately 1% higher mean accuracy throughout the experiments, with final accuracies of

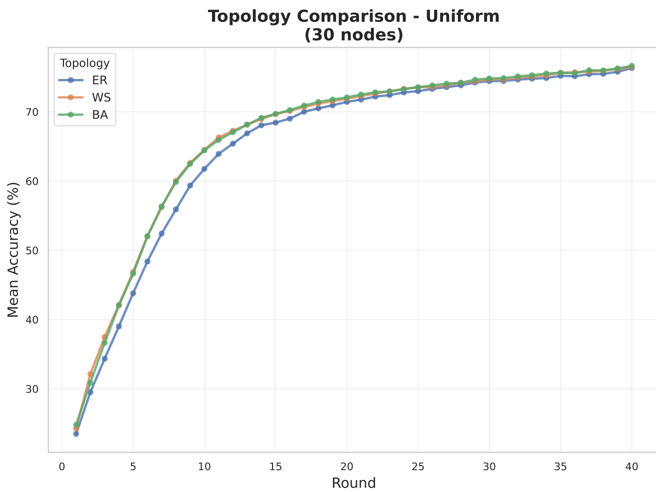

Fig. 2. Average accuracy over rounds using Uniform mixing.

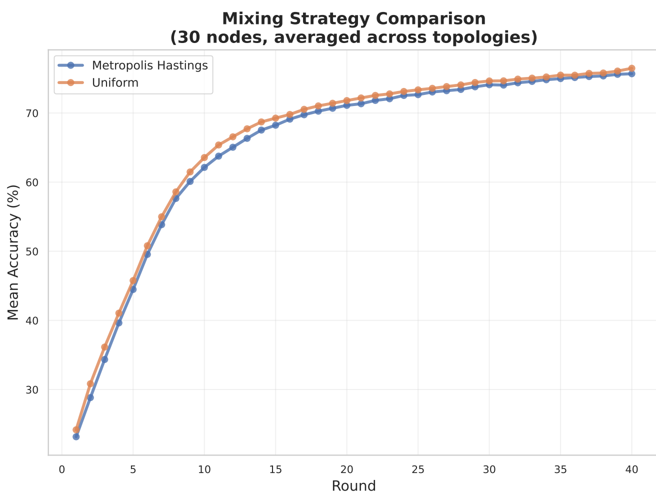

Fig. 3. Comparison of Metropolis-Hastings and Uniform mixing strategies averaged across topologies.

73.6% compared to 72.6% for Metropolis-Hastings. While this difference is relatively modest, Uniform mixing also demonstrates marginally faster convergence in the early rounds (1-10), suggesting more efficient information propagation.

Figure 4 presents all six experimental combinations simultaneously, revealing important interaction effects between topology and mixing strategy. Uniform mixing (shown with dashed lines) consistently outperforms Metropolis-Hastings (solid lines) across all topologies. Notably, BA topology exhibits the largest improvement when switching from Metropolis-Hastings to Uniform mixing, with both convergence speed increasing by 4 rounds and final accuracy improving by 1.83%. ER topology demonstrates the most consistent performance regardless of mixing strategy, while WS performance falls between ER and BA. This suggests that the choice of mixing strategy is particularly important for scale-free (BA) networks due to their heterogeneous degree distribution.

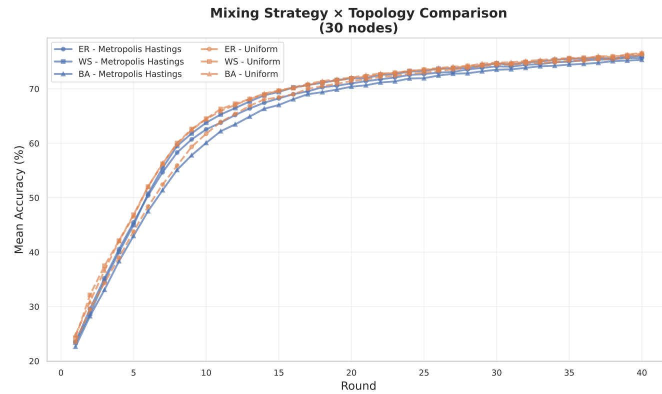

Fig. 4. Detailed comparison of all topology and mixing strategy combinations.

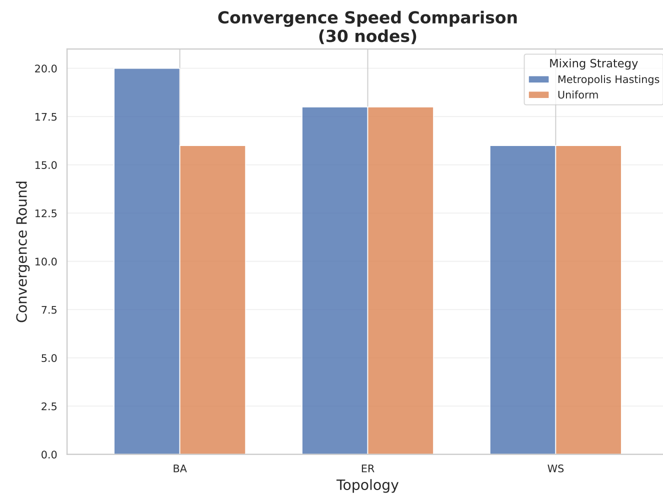

Fig. 5. Convergence speed comparison across topology and mixing strategy combinations.

Figure 5 provides a comparison of convergence rounds required for each topology-mixing combination. WS topology achieves the fastest convergence at 16 rounds regardless of mixing strategy, likely due to its small-world properties which balance local clustering with short average path lengths. The most striking difference appears in BA topology, where Metropolis-Hastings requires 20 rounds to converge while Uniform mixing achieves convergence in only 16 rounds—a 20% reduction. This demonstrates that Uniform mixing significantly reduces the convergence time variability across different topologies, making it a more predictable choice for gossip learning systems.

Figure 6 compares the final mean accuracy achieved by each topology-mixing pair. Uniform mixing achieves higher accuracy across all topologies, with ER-Uniform reaching the highest performance at 74.11%. The accuracy range across all combinations is relatively tight (71.90% to 74.11%, a 2.21% spread), indicating that all configurations reach acceptable performance levels. Comparing Figures 5 and 6 reveals that WS topology offers an optimal trade-off: it achieves fast con-

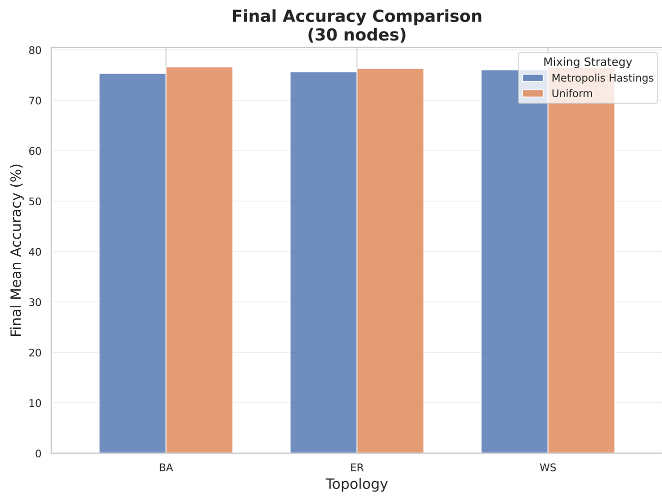

Fig. 6. Final mean accuracy comparison across topology and mixing strategy combinations.

vergence (16 rounds) while maintaining high accuracy (73.50-74.02%). This makes WS networks particularly attractive for gossip learning applications where both speed and accuracy matter.

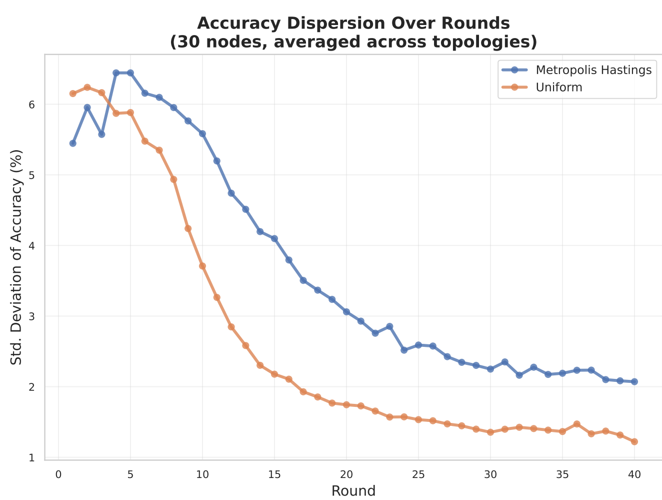

Fig. 7. Average accuracy dispersion over rounds.

Figure 7 displays the accuracy dispersion over all 40 rounds averaged across all topologies. This metric represents the variance in accuracy among all nodes, with lower deviation indicating better consensus and fairness in the distributed learning system. As expected, the standard deviation decreases as rounds progress from over 6% initially to approximately 2.5% for Metropolis-Hastings and 1.5% for Uniform mixing by round 40. This trend reflects successful model weight propagation through the network as gossip communication allows nodes to converge toward similar model parameters. The initial high variance stems from the non-IID data distribution ($\alpha = 1.0$), where each node trains on different local data distributions. When comparing the two mixing strategies, Uniform consistently maintains approximately 1% lower standard deviation throughout training, indicating that it achieves better consensus among nodes. This improved fairness, combined with the higher accuracy and faster convergence shown in previous figures, establishes Uniform mixing as the superior choice for these experimental conditions.

VIII. Conclusions

This work examined how communication topology influences convergence behavior in gossip-based decentralized learning systems under non-IID data distributions. By matching the average degree across Erdős–Rényi, Watts–Strogatz, and Barabási–Albert graphs, we isolated the effects of structural properties such as clustering, path length, and degree heterogeneity from raw connectivity.

Our experimental results demonstrate that, when average degree is held constant, all three topologies achieve comparable final accuracies on Fashion-MNIST, converging within a narrow range of approximately 72–74%. However, differences emerge in convergence speed and stability. Watts–Strogatz networks consistently converged fastest, likely due to their small-world structure combining high clustering with short average path lengths. Erdős–Rényi graphs exhibited stable and predictable convergence behavior, while Barabási–Albert graphs showed greater sensitivity to the choice of mixing strategy due to their heterogeneous degree distribution.

Preliminary observations suggest that topologies with short average path lengths and higher clustering tend to reach consensus more rapidly, while scale-free networks may benefit from hub nodes that accelerate information diffusion. These findings highlight the importance of topology-aware design choices in decentralized learning systems.

Future work will extend this study to larger-scale networks, such as those with 50 to 100 clients, to further evaluate scalability, robustness, and communication efficiency. Increasing the number of nodes will allow for deeper analysis of how graph sparsity, degree heterogeneity, and information diffusion interact at scale.

Additional directions include exploring adaptive and time-varying mixing strategies, dynamic graph topologies, and more expressive model architectures such as convolutional neural networks. Together, these extensions would provide a more comprehensive understanding of how topology-aware design choices impact decentralized learning systems in realistic deployments.

EndNotes

[1]In general topology structure itself doesn't seem to matter but instead the distance between nodes and level of interconnectivity.

[2]https://github.com/KaungMyatNaing9/Topology-Influence-on-Gossip-Learning-Convergence

Author's Contribution

ML designed and formulated the algorithm for gossip learning, did abstract, introduction, and related works as well. ZC split the code from a jupyter notebook into individual python files, ran the experiments on a chameleon node, and wrote the results section. KMN trained the baseline model, developed some graph visualization code, and wrote the system model description and formal problem definition sections. KS extended and refined the gossip learning framework, contributed to the experimental design and methodology, implemented non-IID data partitioning and convergence criteria, and wrote the system setup, methodology, and conclusion sections. All authors read and approved the final manuscript.

References

[1] I. Hegedűs, G. Danner, and M. Jelasity, "Gossip learning as a decentralized alternative to federated learning," *Distributed Applications and Interoperable Systems (DAIS 2019)*, vol. 11534, pp. 74–90, 2019.

[2] I. Hegedűs, G. Danner, and M. Jelasity, "Decentralized learning works: An empirical comparison of gossip learning and federated learning," *Journal of Parallel and Distributed Computing*, vol. 148, pp. 109–124, 2021.

[3] I. Hegedűs, G. Danner, and M. Jelasity, "Decentralized recommendation based on matrix factorization: A comparison of gossip and federated learning," *In Machine Learning and Knowledge Discovery in Database (ECML PKDD 2019)*, vol. 1167, pp. 317–332, 2020.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, , and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," *In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 54, pp. 1273–1282, 2017.

[5] R. Ormándi, I. Hegedűs, and M. Jelasity, "Gossip learning with linear models on fully distributed data," *Concurrency and Computation Practice and Experience*, vol. 25(4), pp. 556–571, 2013.

[6] C. P. Robert, "The metropolis-hastings algorithm," *arXiv preprint arXiv*, vol. 1504.01896, p. 15, 2015.

[7] C. Andrieu and J. Thoms, "A tutorial on adaptive mcmc," *Statistics and Computing*, vol. 18(4), pp. 343–373, 2008.

[8] M. Jelasity, G. R. K. A. M. Voulgaris, S., and M. van Steen, "Gossip-based peer sampling," *ACM Transactions on Computer Systems*, vol. 25(3), p. 8, 2007.

[9] M. Ben-Or and E. Upfal, "Understanding the power of randomization in distributed systems," *ACM Symposium on Theory of Computing (STOC)*, vol. 16, pp. 345–351, 1984.

[10] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.

[11] P. Erdős and A. Rényi, "On random graphs i," *Publicationes Mathematicae*, vol. 6, pp. 290–297, 1959.

[12] Watts, D. J., and S. H. Strogatz, "Collective dynamics of 'small-world' networks.," *Nature*, vol. 393, pp. 440–442, 1998.

[13] Barabási, A.-L., and R. Albert, "Emergence of scaling in random networks.," *Science*, vol. 286(5439), pp. 509–512, 1999.

[14] K. A. Zweig, "Random graphs and network models," *In Networks Analysis Literacy*, vol. Springer, p. 149–181, 2016.

[15] I. Hegedűs, Berta, L. Kocsis, A. A. Benczúr, and M. Jelasity, "Robust decentralized low-rank matrix decomposition," *ACM Transactions on Intelligent Systems and Technology*, vol. 7(4), pp. 62:1 – 62:24, 2016.